Spectrogram Factorization Using Phase Information

R. Mitchell Parry
GVU Center / College of Computing
Georgia Institute of Technology
Atlanta, GA 30332
parry@cc.gatech.edu

Irfan Essa
GVU Center / College of Computing
Georgia Institute of Technology
Atlanta, GA 30332
irfan@cc.gatech.edu

1 Introduction

Spectrogram factorization methods have been proposed for single channel source separation [1–4], audio analysis [5–8] and more recently multichannel source separation [9–11]. All spectrogram factorization approaches incorrectly assume that the mixture spectrogram is the sum of the source spectrograms. In fact, the mixture spectrogram depends on the source spectrograms *and* the phase difference between them. This paper investigates the role of phase in determining the mixture spectrogram and incorporates a probabilistic representation of phase into a novel method for source spectrogram estimation.

When multiple mixture signals are available, independent component analysis (ICA) is a statistical technique that separates as many independent source signals as there are mixture signals. When there is only one mixture signal, the signal may be transformed into a time-frequency representation such as the magnitude of the short-time Fourier transform (*i.e.,* spectrogram). Casey and Westner [1] originated the idea of spectrogram factorization by applying ICA to the single mixture spectrogram, treating each frequency channel as a separate mixture signal. Using this approach, ICA separates as many sources as frequency channels. However, the expressiveness of each source is necessarily diminished. Each source spectrogram is a rank-one matrix formed by the product of a column vector containing the spectral shape and a row vector containing the time-varying gain. The actual source spectra are deemed to be a combination of multiple rank-one source spectrograms.

The problem with ICA for spectrogram factorization is that it extracts components that have negative elements, whereas spectrogram data is always non-negative. Therefore, non-negative matrix factorization (NMF) has been proposed for source spectrogram estimation. NMF does not require independence but maintains non-negative elements. An underlying assumption of ICA- and NMF-based approaches is that the mixture spectrogram is the sum of the source spectrograms. This assumption is valid only in the unlikely event that all sources have the *same* phase at *every* time-frequency point or in the trivial case when only one source is active. In all other cases, the mixture spectrogram also depends on the phase information in the short-time Fourier transform (STFT) of the sources. We present a method to incorporate the unknown source phase information into the estimation of the source spectrograms using a probabilistic representation of phase.

2 Spectrogram factorization

The first step for spectrogram factorization methods is to convert the mixture signal into a time-frequency representation such as the complex-valued short-time Fourier transform (STFT). Because the STFT contains the phase information for each source, the mixture STFT can be constructed precisely as the sum of the source STFT matrices. However, this model does not extend to the spectrogram (*i.e.,* the absolute value of each element of an STFT matrix). This is due to the nonlinearity

of the absolute value function and is analogous to the following inequality.

$$|a+b| \neq |a| + |b| \tag{1}$$

In contrast, spectrogram factorization techniques typically assume that the *magnitude of the sum* of source STFTs (*i.e.,* mixture spectrogram) is equal to the *sum of the magnitudes* of the source STFTs (*i.e.,* source spectrograms). Non-negative matrix factorization estimates the mixture spectrogram, $\mathbf{V}$, as the sum of multiple source component spectrograms, $\mathbf{C}_r$:

$$\mathbf{V} \approx \mathbf{W}\mathbf{H}^T = \sum_r \mathbf{w}_r \mathbf{h}_r^T = \sum_r \mathbf{C}_r \tag{2}$$

where the column vector $\mathbf{w}_r$ and $\mathbf{h}_r$ represents the spectral shape and amplitude envelope of the r-th source component, respectively. Non-negative matrix factorization estimates $\mathbf{W}$ and $\mathbf{H}$ by minimizing a distance metric such as the squared Euclidian distance [12]:

$$\|\mathbf{V} - \mathbf{W}\mathbf{H}^T\|^2 = \sum_{kt} \left(\mathbf{V}_{kt} - \left[\mathbf{W}\mathbf{H}^T\right]_{kt}\right)^2 = \sum_{kt} \left(\mathbf{V}_{kt} - \sum_r \left[\mathbf{C}_r\right]_{kt}\right)^2 \tag{3}$$

This approach optimizes one possible configuration of phase, namely when the sources have equal phase or when only one source is active. By incorporating the true distribution of phase, we improve the estimates of $\mathbf{W}$ and $\mathbf{H}$.

3 Probabilistic representation of phase

By considering the phase at each time-frequency point of each source to be a uniformly distributed random variable, we derive the probability density function of the mixture spectrogram given the source spectrograms. For the case of two components at one time-frequency point, the magnitude of the mixture, $v = \mathbf{V}_{kt}$, is a function of the magnitude of each source component, $c_r = [\mathbf{C}_r]_{kt}$ and the phase difference between them:

$$v = \sqrt{c_1^2 + c_2^2 + 2c_1c_2\cos\theta} \tag{4}$$

Because of the circularity of phase, the difference in two uniformly distributed random phases is also a uniformly distributed random variable, $\theta = U(-\pi, \pi)$. This allows us to derive the likelihood of v given c_1 and c_2:

$$p(v|c_1, c_2) = \frac{2v}{\pi\sqrt{-(v+c_1+c_2)(v+c_1-c_2)(v-c_1+c_2)(v-c_1-c_2)}} \tag{5}$$

By using uninformative priors on c_1 and c_2, we approximate $p(c_1, c_2|v) \propto p(v|c_1, c_2)$. We propose maximizing this equation with respect to c_1 and c_2 by minimizing the following:

$$\underset{c_1,c_2}{\operatorname{argmin}} \left((v+c_1-c_2)(v-c_1+c_2)(v-c_1-c_2)/v^2\right)^2 \tag{6}$$

which reaches a minimum at the maxima of Equation 5. Notice that in terms of v, c_1 and c_2, the standard NMF solution in Equation 2 is the following:

$$\underset{c_1,c_2}{\operatorname{argmin}} \left(v - c_1 - c_2\right)^2 \tag{7}$$

Figure 1 illustrates Equation 5 as a function of c_1 and c_2 for $v = 1$. The standard NMF solution minimizes the distance to the line $v = c_1 + c_2$, whereas the true distribution has energy along a corridor defined by the asymptotes $v = c_1 - c_2$ and $v = c_2 - c_1$.

We tested our approach over 1000 trials using random $\mathbf{W}$ and $\mathbf{H}$ matrices and uniformly distributed phase. As compared to the standard NMF solution, our approach provided a 28% improvement in the mean square error of the estimated $\mathbf{W}$ and $\mathbf{H}$ matrices. For comparison, Figure 2 shows the scatter plots for one representative trial. Using Equation 6 we attain a distribution that more closely resembles Figure 1 evidenced by data points clustering along $v = c_1 - c_2$ and $v = c_2 - c_1$.

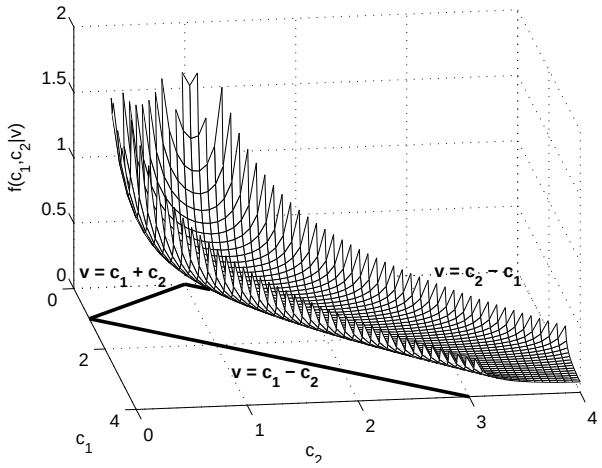


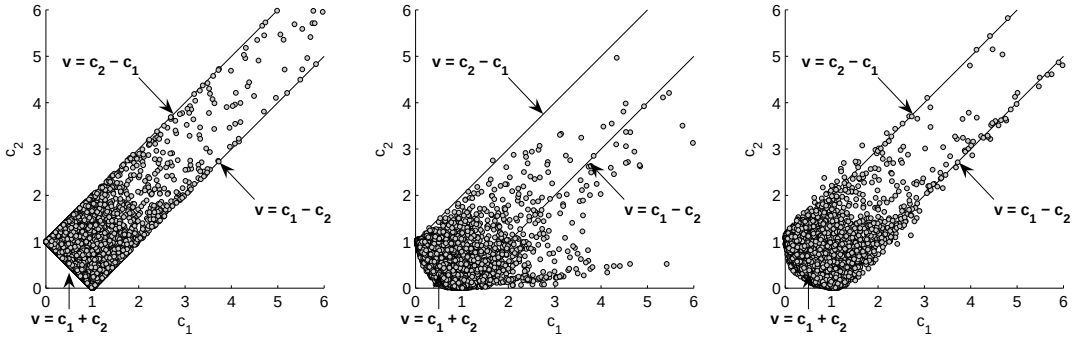
Figure 1: Energy function for a and b when $v = 1$.

4 Conclusion and future work

We have shown that phase plays an important role in the determination of the mixture spectrogram from a number of source spectrograms. By incorporating a probabilistic representation of phase, we propose an improvement on NMF that more closely follows the true distribution of source spectrogram points given the mixture spectrogram. However, extending this analytical solution to more than two components presents quite a challenge. In general, for R components Equation 4 becomes the following:

$$v = \sqrt{\sum_{i=\{1..R\}} c_i^2 + \sum_{i,j=\{1..R\}} c_i c_j \cos\theta_{ij}} \tag{8}$$

where θ_{ij} is the phase difference between component i and j. This leads to $R-1$ independent and identically distributed random variables, $\{\theta_{1j}|j = 2..R\}$, with the remaining dependent variables determined by: $\theta_{ij} = \theta_{1j} - \theta_{1i}$. Deriving the likelihood of the mixture given the source components requires integration over the independent variables where the domain of integration is nontrivial. Therefore, we are exploring numerical and sampling based approaches. By examining a histogram of $p(v|\{c_i\})$ constructed from a sufficiently large sample of points, we have observed that it has asymptotes (*i.e.,* spikes) at positive values of $v = \sum_i \pm c_i$. In addition, the shape of the distribution moves from a U-shaped "trough" for $R = 2$ to a more bell-shaped distribution for larger R. We intend to leverage this information to apply our technique to greater numbers of components.



(a) Correct (b) NMF (c) Our approach

Figure 2: Scatter plot of bins for one representative trial.

References

[1] M. Casey and W. Westner. Separation of mixed audio sources by independent subspace analysis. In *Proceedings of the International Computer Music Conference*, Berlin, August 2000.

[2] P. Smaragdis. *Redundancy Reduction for Computational Audition, a Unifying Approach*. PhD thesis, MAS Department, Massachusetts Institute of Technology, 2001.

[3] T. Virtanen. Separation of sound sources by convolutive sparse coding. In *ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio Processing*, 2004.

[4] B. Wang and M. D. Plumbley. Investigating single-channel audio source separation methods based on non-negative matrix factorization. In *ICA Research Network International Workshop*, pages 17–20, September 2006.

[5] S. A. Abdallah and M. D. Plumbley. Polyphonic transcription by non-negative sparse coding of power spectra. In *Proceedings of the International Conference on Music Information Retrieval*, pages 318–325, Barcelona, Spain, October 2004.

[6] D. FitzGerald, E. Coyle, and B. Laylor. Sub-band independent subspace analysis for drum transcription. In *Proceedings of International Conference on Digital Audio Effects*, pages 65–69, Hamburg, Germany, September 2002.

[7] P. D. O'Grady and B. A. Pearlmutter. Convolutive non-negative matrix factorisation with sparseness constraint. In *Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing*, September 2006.

[8] P. Smaragdis and J. C. Brown. Non-negative matrix factorization for polyphonic music transcription. In *Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 177–180, New Paltz, NY, October 2003.

[9] D. FitzGerald, M. Cranitch, and E. Coyle. Non-negative tensor factorisation for sound source separation. In *Proceedings of Irish Signals and Systems Conference*, Dublin, Ireland, September 2005.

[10] R. M. Parry and I. Essa. Estimating the spatial position of spectral components in audio. In *Proceedings of International Conference on Independent Component Analysis and Blind Signal Separation*, pages 666–673, Charleston, SC, March 2006.

[11] D. FitzGerald, M. Cranitch, and E. Coyle. Sound source separation using shifted non-negative tensor factorisation. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Toulouse, France, May 2006.

[12] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems 13*, pages 556–562. MIT Press, 2001.